

Portafogli vs S&P 500

Progetto di Financial Markets Analytics

Potertì Daniele 844892

Sanvito Alessio 844785

Sanvito Simone 844794

Sommario

Introduzione	3
Scaricamento dati	3
Preprocessing	4
Costruzione modello Rolling	5
Assegnazione dei pesi agli stock nel portfolio	6
Rebalancing settimanale e calcolo dei ritorni del portfolio	7
Risultati finali - confronto Portfolio vs Market Index	8
Conclusioni	20

1. Introduzione

L'obiettivo di questo progetto è stato quello di comprendere il meglio possibile le caratteristiche strutturali che il rischio comporta negli investimenti in portafogli reali.

Per fare ciò sono stati considerati i dati storici di alcuni degli asset presenti all'interno di un'indice di mercato e, dopo aver calcolato i ritorni logaritmici e aver eseguito una regressione rolling su di essi, sono stati costruiti diversi portfoli secondo diverse strategie.

Per selezionare gli stock da inserire all'interno di questo portfolio sono stati calcolati diverse misure quali R^2 , σ^2 , β^2 , σ^2_M , $\sigma^2_{e_i}$, α_i , r_i ; dopo aver calcolato il valore di queste misure per ogni stock è stato selezionato il 2% e 10% di titoli ordinandoli secondo una particolare combinazione.

Sono stati costruiti diversi equally weighted portfolios composti da n stock con pesi uguali ($1/n$ per ogni titolo); è stato effettuato un rebalancing settimanale e ogni volta sono stati calcolati i ritorni tramite la media pesata dei ritorni degli stock.

Infine sono stati paragonati i rendimenti dei portfoli con l'indice di mercato attraverso un grafico per comprendere al meglio le differenze di performance.

2. Scaricamento dati

I dati utilizzati per la realizzazione del progetto sono stati scaricati da Bloomberg, tramite il terminale BLOOMBERG ANYWHERE with Bloomberg Excel Add-In. Questo terminale ha permesso di utilizzare una versione remota di Excel facendola eseguire sui server di Bloomberg, interagendo con gli API data dell'azienda.

È stato così effettuato il download dei dati, con granularità giornaliera, relativi ai last price delle aziende presenti all'interno dello Standard & Poor's 500.

Lo Standard & Poor's 500 è il più importante indice azionario nordamericano. Sebbene storicamente siano nati prima gli indici Dow Jones, questo paniere ha assunto con il tempo maggiore importanza presso gli investitori. È infatti il principale benchmark azionario relativo ai titoli quotati a Wall Street e fornisce un incredibilmente ampio ventaglio di prodotti derivati, quali futures, opzioni e certificati.

Lo S&P 500 contiene 500 titoli azionari di altrettante società quotate a New York (NYSE e Nasdaq), rappresentative dell'80% circa della capitalizzazione di mercato, che vengono selezionate da un apposito comitato.

Successivamente si è proceduto con il download dei dati relativi all'indice di mercato: sempre tramite il terminale di Bloomberg, sono stati scaricati i `px_last` (last prices) di ogni giorno dell'indice S&P500.

Sia per gli stock che per l'indice di mercato è stata presa come data di inizio della time series il giorno 2012-03-16. I dati sono stati scaricati da questa data in avanti, per un periodo decennale, fino al 2022-06-14, tramite la formattazione 5D-5 BUS NO HOLIDAY, per prendere i 5 giorni lavorativi (lun-ven).

In particolare, per scaricare i dati relativi agli stock è stata necessaria una ulteriore fase: sono stati selezionati tutti i titoli che, nel periodo in esame, sono stati componenti dell'indice S&P500; partendo dalla configurazione attuale di questo indice, sono stati analizzati tutti gli ingressi e tutte le uscite dei titoli dall'indice (https://en.wikipedia.org/wiki/List_of_S%26P_500_companies) e si è creato un dataset che raccoglie tutti i titoli che hanno fatto parte dello S&P 500 da settembre 2017 in poi. In questo modo è stato possibile evitare di avere dei dati affetti da bias, più precisamente dal survivorship bias (che è un tipo di bias legato alla “sample selection” in cui non vengono considerate le osservazioni che hanno già cessato di esistere; rappresenta la tendenza a considerare solo i titoli esistenti o "sopravvissuti" o “vincitori” quando si misura la performance di un portafoglio).

3. Preprocessing

Il dataset scaricato è costruito come segue:

- ogni riga rappresenta un giorno, in particolare vengono considerati inizialmente 2673 giorni, dal 2012-03-16 al 2022-06-14
- ogni colonna rappresenta un titolo appartenente all'indice di mercato S&P500; la prima colonna Unnamed:0 contiene tutte le date

La prima operazione volta a rendere il dataset utilizzabile è stata quella di reindexing del dataset; infatti, la colonna Unnamed:0 è stata settata come index del dataset per rendere più facile l'interazione con il dataset. In seguito sono stati cambiati i nomi delle colonne in modo da associare ogni colonna relativa a uno stock al suo simbolo (es. Apple → AAPL).

Per tutte le task relative al progetto, sono stati considerati i dati a partire da gennaio 2017 in poi.

In seguito, i dati importati sono stati modificati manualmente, facendo in modo che rispecchiassero gli aggiornamenti della composizione dell'indice durante gli anni considerati; per questa fase sono stati considerati i dati traslati di 180 giorni in avanti rispetto a gennaio 2017, ovvero a partire da settembre 2017.

In questo scenario si possono verificare due situazioni:

- uno stock esce dall'indice: dal giorno di uscita dello stock dall'indice fino alla fine del periodo (o fino all'eventuale data di reinserimento dello stock nell'indice) è stato inserito, al posto del prezzo relativo a quei giorni, il valore N/A;
- uno stock entra nell'indice: in tutto il periodo precedente al giorno di ingresso dello stock nell'indice (fino al giorno in cui effettivamente è stato aggiunto) è stato inserito, al posto del prezzo relativo a quei giorni, il valore N/A (facendo attenzione anche in questo caso che fosse la prima apparenza all'interno di S&P500).

Il data frame è stato così salvato e utilizzato per le fasi successive.

4. Costruzione modello Rolling

In questa fase sono stati calcolati (e salvati negli appositi data frame) i log returns dei vari stock e i log return dell'indice di mercato.

Ci si concentra sui log returns perché eliminano le proprietà non stazionarie del dataset, rendendo i dati finanziari più stabili. I log returns, ovvero i valori ottenuti dal logaritmo naturale del simple gross return di un asset, sono stati calcolati come segue:

$$r_t = \ln(1 + R_t) = \ln \frac{P_t}{P_{t-1}} = p_t - p_{t-1} \quad \text{where} \quad p_t = \ln P_t$$

In particolare, è stato calcolato il logaritmo del rapporto tra il prezzo al tempo t ed il prezzo al tempo $t-1$ (ovvero la variabile shiftata di 1, in quanto facendo lo shift si prende il valore, di quella colonna, contenuto nella cella appena sopra).

Anche in questo caso sono stati presi solo i valori relativi agli ultimi 5 anni.

È stato costruito un modello rolling che fa un sample sui primi 180 giorni. Al 180esimo giorno avremo i parametri della regressione relativi ai 180 giorni precedenti, per ogni asset.

In seguito, questo modello, effettua una slide della window di 7 giorni. A quel punto al 187esimo giorno verranno specificati i parametri della regressione relativi ad ogni asset calcolati sui 180 giorni precedenti. Così via fino a raggiungere la fine del dataset, ovvero 5 anni dopo.

Sono stati poi calcolati, per tutte le settimane di riferimento del modello rolling, tramite la regressione dei residui:

- beta (β_i): il calcolo del beta viene utilizzato per aiutare gli investitori a capire se un'azione si muove nella stessa direzione del mercato; fornisce informazioni su quanto sia volatile o rischioso un titolo rispetto al resto del mercato;
- alfa (α_i): eccesso di ritorno; il rendimento in eccesso di un investimento rispetto al rendimento di un indice di riferimento è l'alfa dell'investimento;
- p-value associato ad alfa: sta ad indicare la significatività del valore di alfa che è stato trovato; si cerca un valore di alfa minore o uguale (nel nostro caso) a 0.1, in modo tale da avere un valore statisticamente significativo e quindi stabile nel tempo;
- R^2 : misura il legame tra la variabilità dei dati e la correttezza del modello statistico utilizzato;
- rischio totale (σ_i^2): rischio totale di un asset, è composto dalla somma tra rischio sistematico e rischio specifico;
- rischio sistematico (varianza del mercato, σ_M^2): rappresenta il rischio di mercato; non calcolato tramite la regressione, ma tramite la varianza dei ritorni dell'indice di mercato;

- rischio specifico (varianza dei residui, $\sigma_{e_i}^2$): rappresenta il rischio specifico di un asset, indica la parte dei movimenti, relativi al prezzo, legati a specifici eventi di un singolo titolo;
- ritorni (r_i): ritorni che si ottengono quando si fa una regressione; sono calcolati facendo $\alpha_i + \beta_i * R_M + e_i$ (dove e_i è un proxy del rischio specifico).

Questi parametri sono poi stati utilizzati per prendere le decisioni relative alla costruzione dei diversi portafogli. In particolare, per la costruzione del portafoglio di una specifica settimana, sono stati utilizzati i valori relativi a questi parametri calcolati nella settimana precedente.

5. Assegnazione dei pesi agli stock nel portfolio

Per assegnare i pesi agli stock all'interno del portfolio è stato utilizzato il metodo di misurazione chiamato equal weight che attribuisce la stessa importanza a ciascun titolo in un portafoglio. Per ogni settimana vengono considerati solamente i portafogli che stanno all'interno dell'indice.

Con la procedura equal weight si assegnano dunque pesi uguali agli stock nel portfolio: nel caso dei portfoli in cui si seleziona il 2% del numero totale di stock, ogni stock avrà peso $1/(2\% * \text{numero totale di stock})$, nel caso dei portfoli in cui si seleziona il 10% del numero totale di stock, ogni stock avrà peso $1/(10\% * \text{numero totale di stock})$.

6. Selezione degli indici per la costruzione del portfolio

Sono stati costruiti diversi tipi di portafogli in modo da poterli confrontare per vedere quale di essi ottiene i ritorni e le performance migliori.

In particolare si è scelto di selezionare gli stock per comporre i diversi portafogli in questo modo:

1. ordinando l'r-squared relativo a ogni stock in ordine decrescente sono stati creati quattro portafogli: uno formato dal primo 10% degli stock con r-squared maggiori, uno formato dall'ultimo 10% dei titoli (che corrispondono agli stock con gli r-squared più bassi), uno con il primo 2% degli r-squared maggiori e l'ultimo con il primo 2% dei titoli con r-squared minimi.
 - 1.1. 2% titoli con R-squared massimo
 - 1.2. 2% titoli con R-squared minimo
 - 1.3. 10% titoli con R-squared massimo
 - 1.4. 10% titoli con R-squared minimo
2. Sono stati valutati due parametri, ovvero alpha (termine utilizzato negli investimenti per descrivere la capacità di una strategia di investimento di battere il mercato, o il suo "limite") e la significatività di alpha (alpha è stato considerato significativo se il p-value è minore di 0.1,

indicando così un intervallo di confidenza al 90%); quindi, considerando i titoli con alpha significativo, essi sono stati ordinati in base al valore crescente e decrescente di alpha selezionando i primi valori corrispondenti al 10% e al 2% (creando quindi 4 portafogli: l'upper bound del 10% dei portafogli totali con alpha massimi, l'upper bound del 10% dei portafogli totali con alpha minimi, l'upper bound del 2% dei portafogli totali con alpha massimi, l'upper bound del 2% dei portafogli totali con alpha minimi).

- 2.1. 2% titoli con alpha significativo e valore di alpha massimo
 - 2.2. 2% titoli con alpha significativo e valore di alpha minimo
 - 2.3. 10% titoli con alpha significativo e valore di alpha massimo
 - 2.4. 10% titoli con alpha significativo e valore di alpha minimo
3. Come ultimo indice è stato considerato il rapporto tra ritorno e rischio; anche in questo caso, come nei punti 1 e 2, l'indice è stato ordinato prima in ordine crescente, poi in ordine decrescente e sono stati selezionati il 2% e il 10% sia dell'upper quartile che del lower quartile.
- 3.1. 2% titoli con rapporto ritorno/rischio massimo
 - 3.2. 2% titoli con rapporto ritorno/rischio minimo
 - 3.3. 10% titoli con rapporto ritorno/rischio massimo
 - 3.4. 10% titoli con rapporto ritorno/rischio minimo

7. Rebalancing settimanale e calcolo dei ritorni del portfolio

È stata implementata anche una funzionalità che consente di ribilanciare il portfolio ogni sette giorni; in base al parametro che viene selezionato come indice nella [sezione 6](#) e in base al valore assunto da ogni stock (relativamente a quel parametro) alla fine dei sette giorni, vengono selezionati i titoli che rispettano la condizione (2% o 10%, minimo o massimo) e si ripetono la procedura di assegnazione dei pesi e la simulazione di investimento.

Per calcolare i ritorni di ogni portfolio è stata fatta una media della somma pesata dei ritorni di ogni giorno, usando come numero di stock, il 2% o il 10% di tutti gli stock, in base al portfolio da costruire.

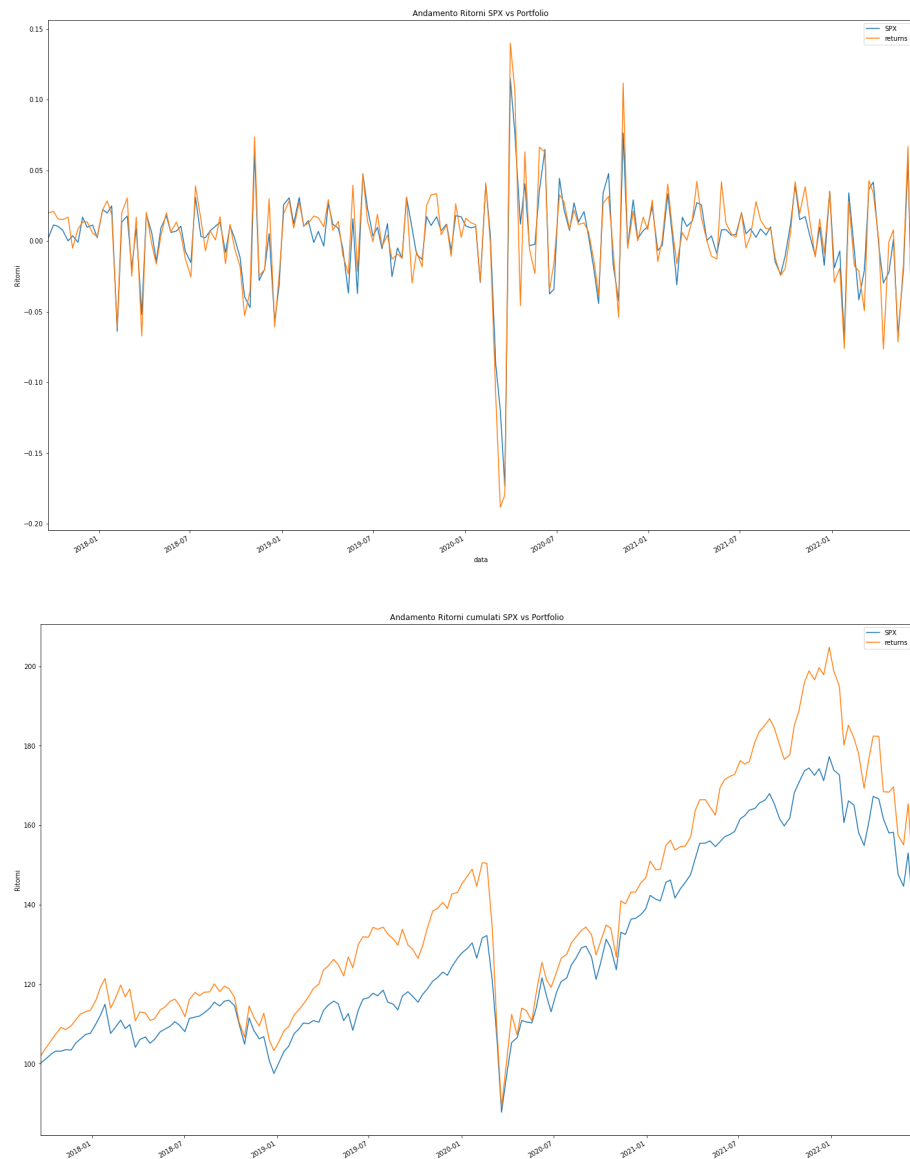
Sono stati anche calcolati i ritorni cumulati, sia di ogni portfolio, che dell'indice di riferimento, in modo da poterli confrontare e analizzare i risultati ottenuti.

Infine, sono state create due tabelle che riportassero il ritorno medio, la volatilità ed il rapporto tra ritorno e rischio di ogni portafoglio e dell'indice di mercato; una è relativa alla selezione del 2% degli stock, l'altra alla selezione del 10%.

8. Risultati finali - confronto Portfolio vs Market Index

1. R-squared

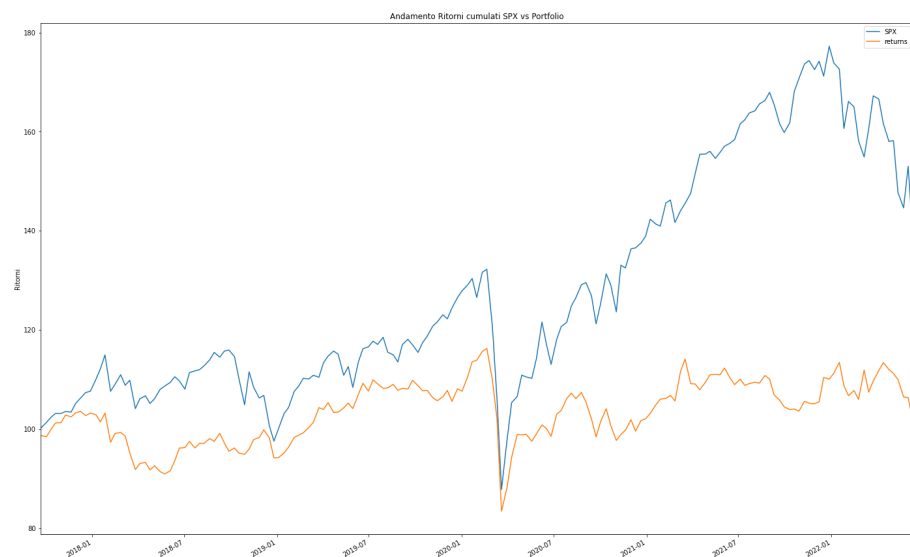
1.1. 2% titoli con R-squared massimo



Nel caso di R-squared massimi, si può notare che il portfolio overperforma il mercato (relativamente ai ritorni) in tutto il periodo temporale; grazie alla

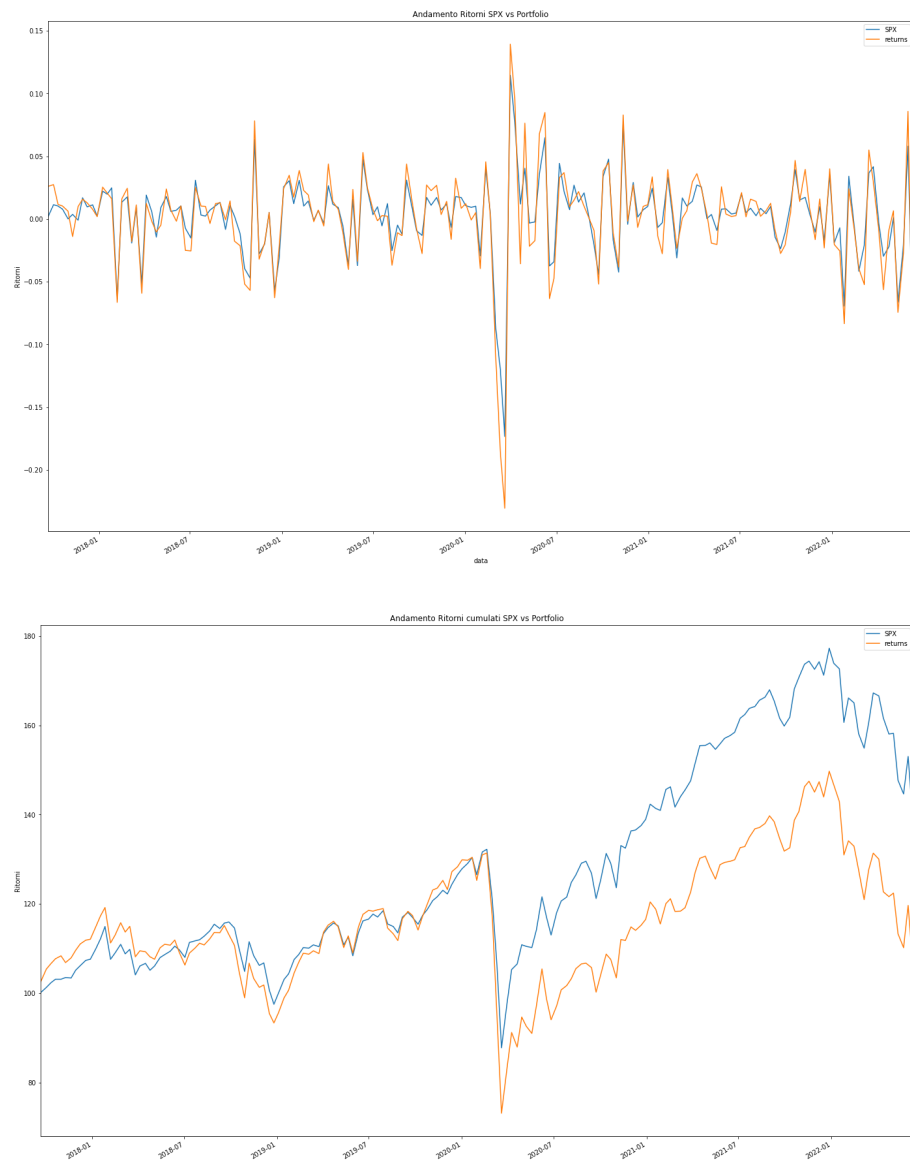
selezione degli R-squared migliori si riesce a creare un portfolio con i titoli che si adattano meglio all'andamento dell'indice di mercato.

1.2. 2% titoli con R-squared minimo



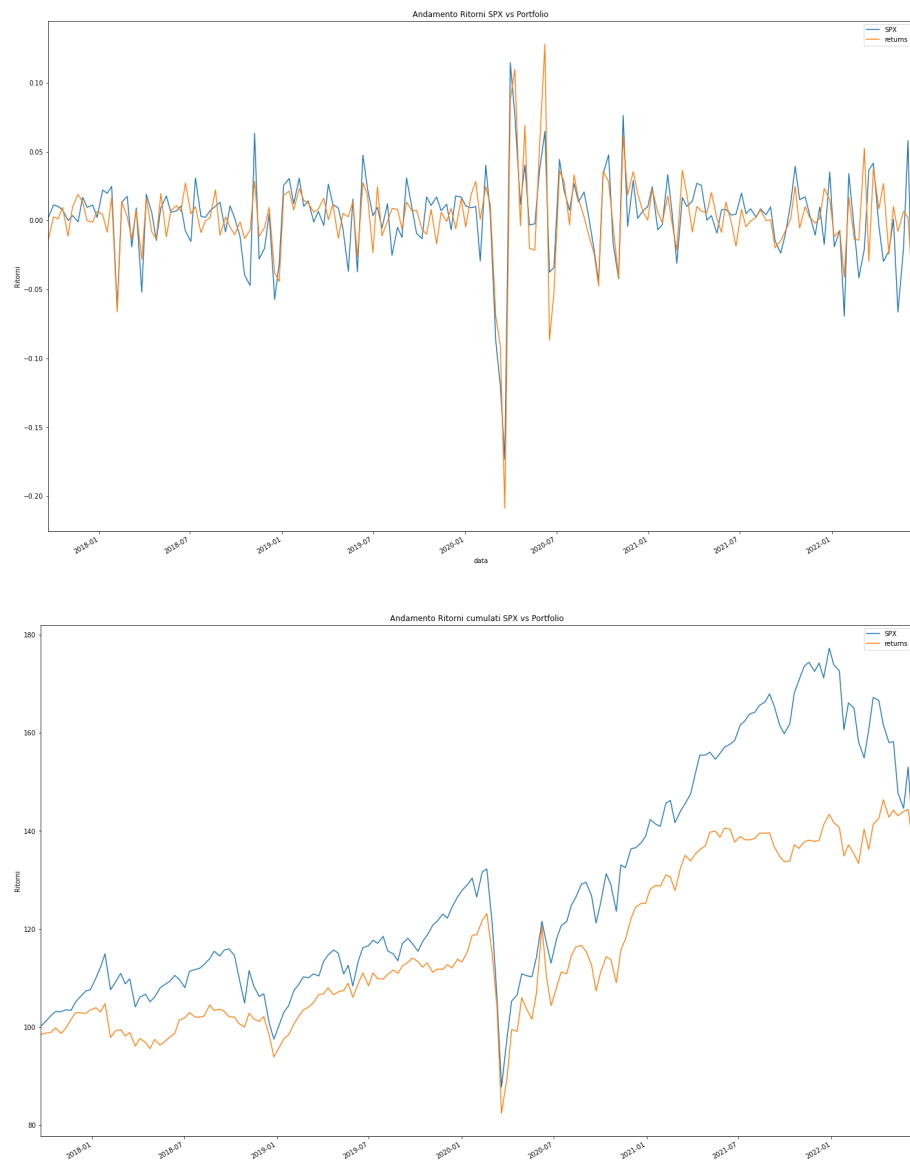
Ragionevolmente scegliendo il lower bound dell'R-squared, il portfolio che si ottiene ha dei ritorni cumulati inferiori rispetto al mercato.

1.3. 10% titoli con R-squared massimo



Aumentando il numero di titoli da inserire all'interno del portfolio, si può notare che non si riesce ad overperformare il mercato; fino a poco prima del picco negativo legato al periodo iniziale del Covid il portfolio è pressoché in linea con l'andamento e i ritorni del mercato, ma successivamente non riesce a ottenere dei ritorni tanto alti quanto quelli del mercato.

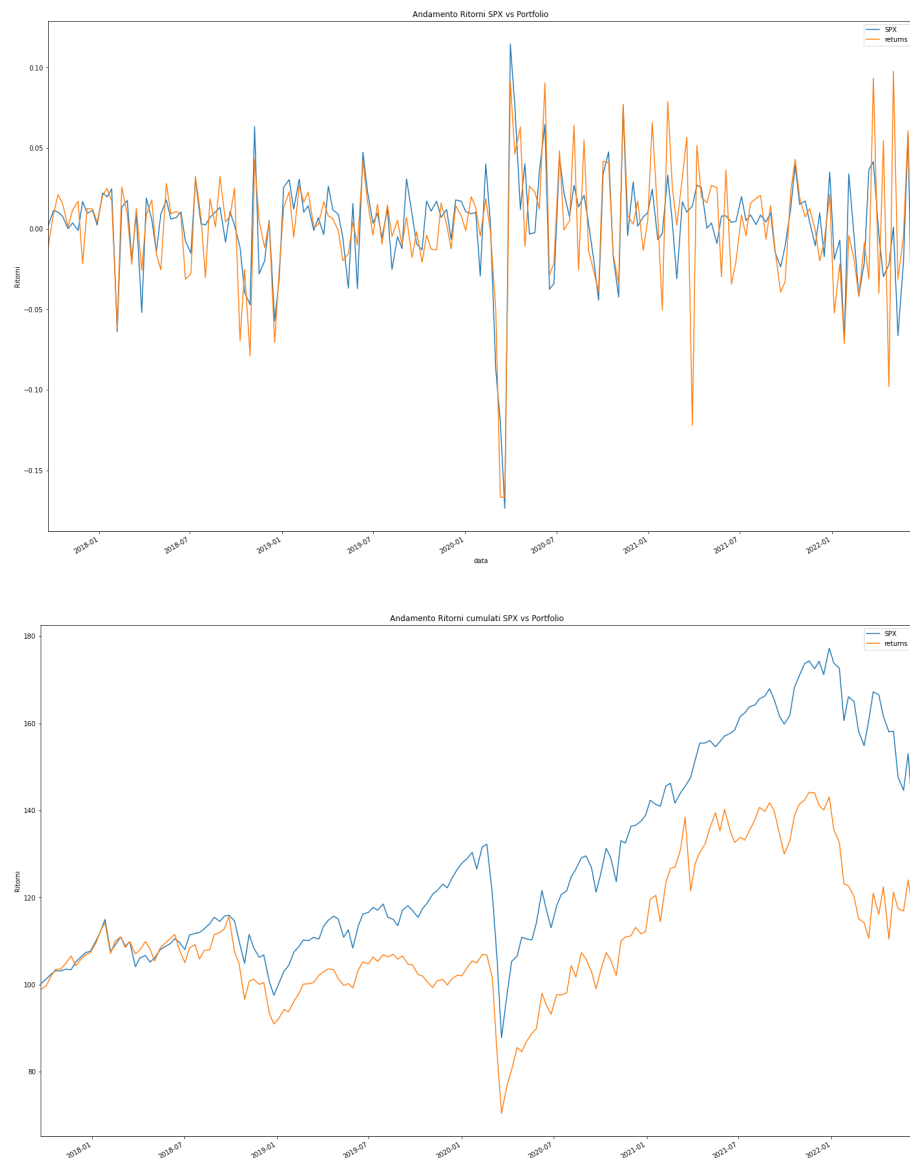
1.4. 10% titoli con R-squared minimo



In questo caso si può notare invece un leggero miglioramento rispetto al caso analogo (R-squared minimi), nel senso che i ritorni del portfolio sembrano più cointegrati con quelli di S&P500 (nonostante siano inferiori in tutto il periodo di investimento). Alla fine del periodo i ritorni del portfolio sono molto vicini a quelli del mercato.

2. Alpha

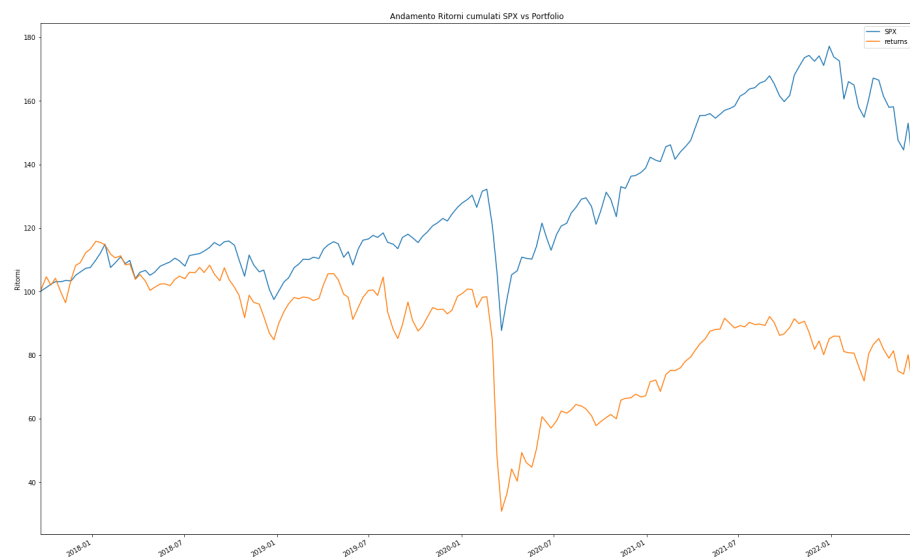
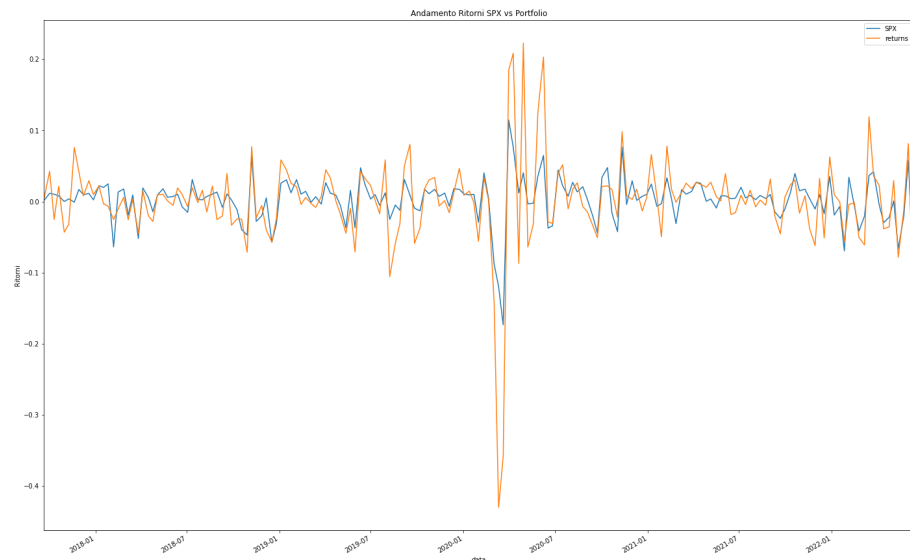
2.1. 2% titoli con alpha significativo e valore di alpha massimo



Il portfolio nelle primissime settimane riesce a overperformare il mercato in termini di ritorni; questo comportamento è in linea con il significato di alpha. Con il passare del tempo, nonostante siano state scelte le alpha significative (quindi quelle che dovrebbero essere più stabili nel tempo), si può notare

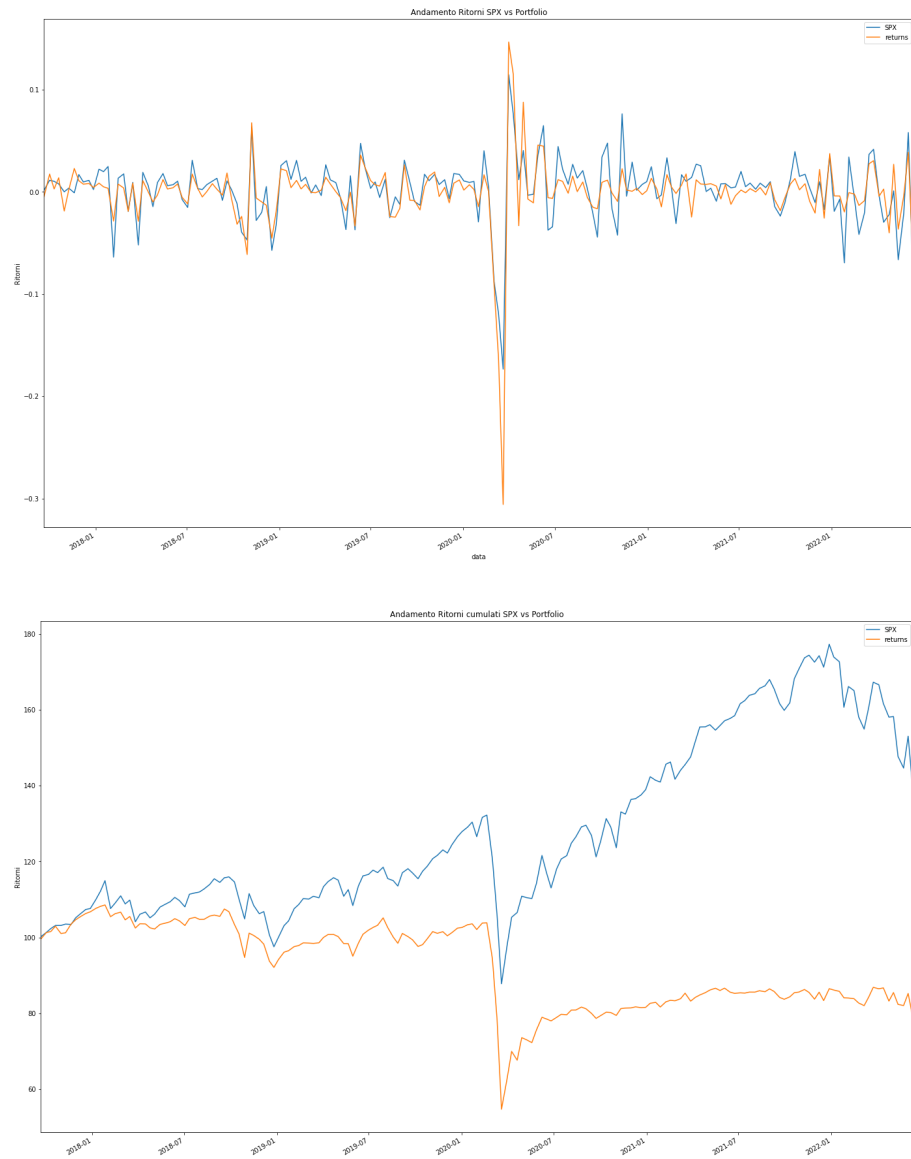
come il portfolio creato non riesca a ottenere dei ritorni tanto grandi quanto quelli del portfolio.

2.2. 2% titoli con alpha significativo e valore di alpha minimo



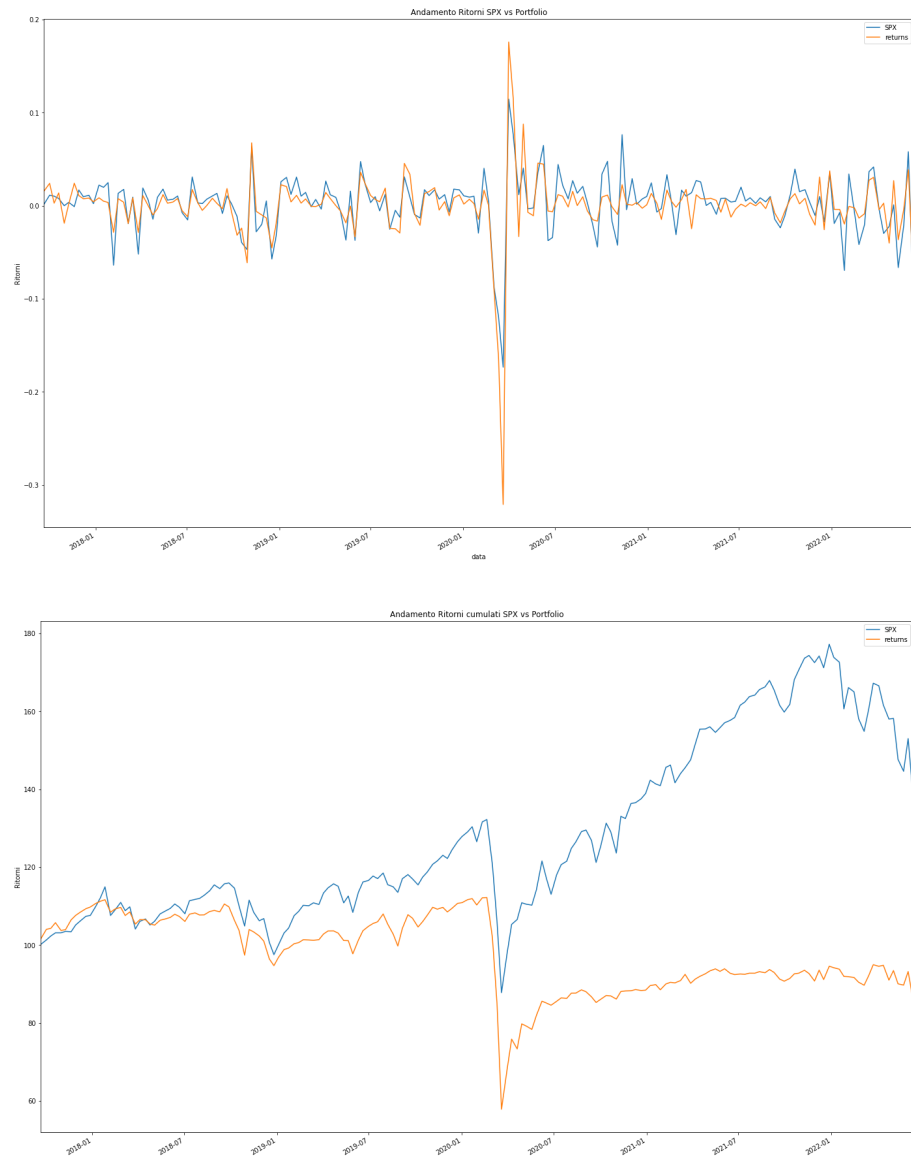
Scegliendo di inserire gli stock con valore minimo di alpha è evidente come il periodo di overperformance sia ancora più ridotto rispetto al caso precedente; si ottengono dei ritorni che alla fine del periodo sono circa la metà di quelli ottenuti dall'indice di mercato.

2.3. 10% titoli con alpha significativo e valore di alpha massimo



Aumentando il numero di stock viene eliminata anche la componente di overperformance iniziale.

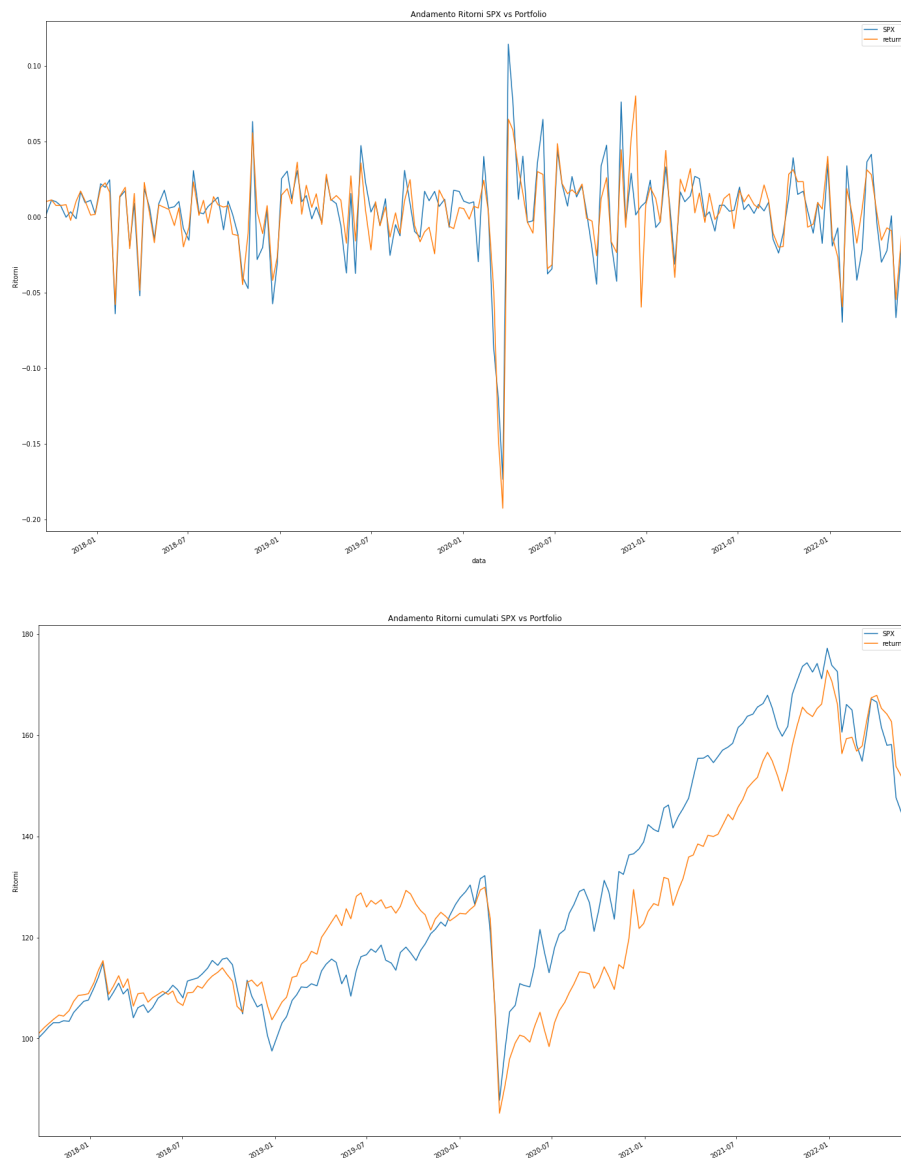
2.4. 10% titoli con alpha significativo e valore di alpha minimo



Nel caso in cui viene scelto il 10% di stock con alpha minore si registra un breve periodo di overperformance, ma nel complesso i ritorni ottenuti sono minori rispetto al caso con il 2% degli stock con alpha minimo.

3. Ritorno/rischio

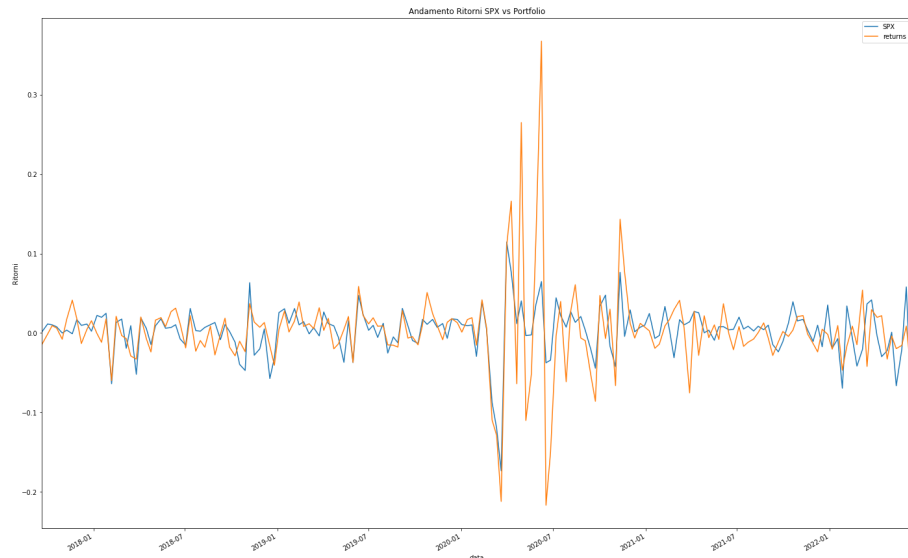
3.1. 2% titoli con rapporto ritorno/rischio massimo



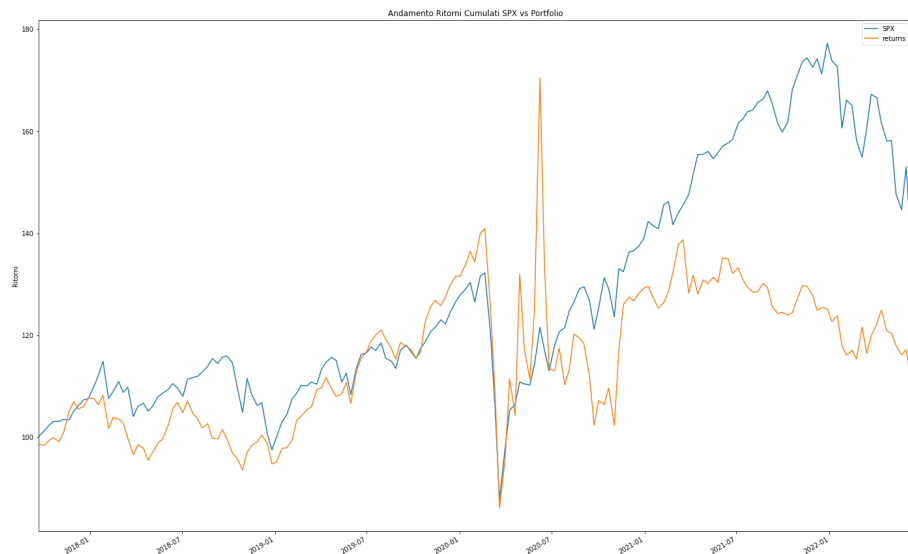
Il portafoglio costruito con il massimo rapporto ritorno-rischio possibile si muove praticamente con il mercato nel primo periodo e ottiene ritorni migliori tra fine 2018 e inizio 2020; poi, con la crisi dovuta al Covid, ha un picco negativo (come succede anche al mercato) e performa sempre peggio rispetto al mercato, fino agli ultimi mesi (nel 2022 circa) in cui torna a muoversi con il

mercato (oscillando tra ritorni migliori di quelli del mercato e ritorno di poco peggiori).

3.2. 2% titoli con rapporto ritorno/rischio minimo



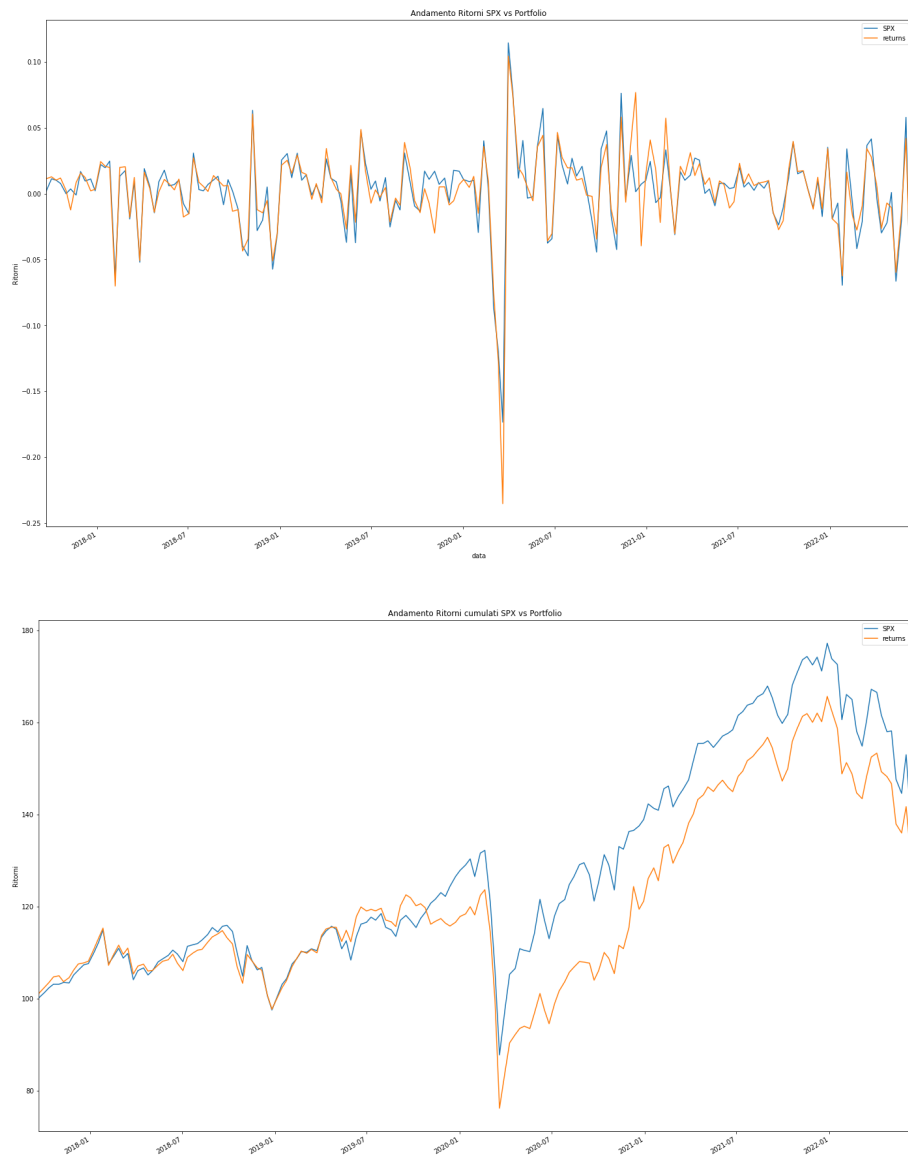
Grande volatilità del portafoglio rispetto al mercato nel periodo centrale.



Considerando il 2% degli stock con il rapporto rischio ritorno minimo si può notare che generalmente il portafoglio underperforma, ottiene dei ritorni minori rispetto a quelli del mercato, fino a luglio 2019; successivamente per tutto

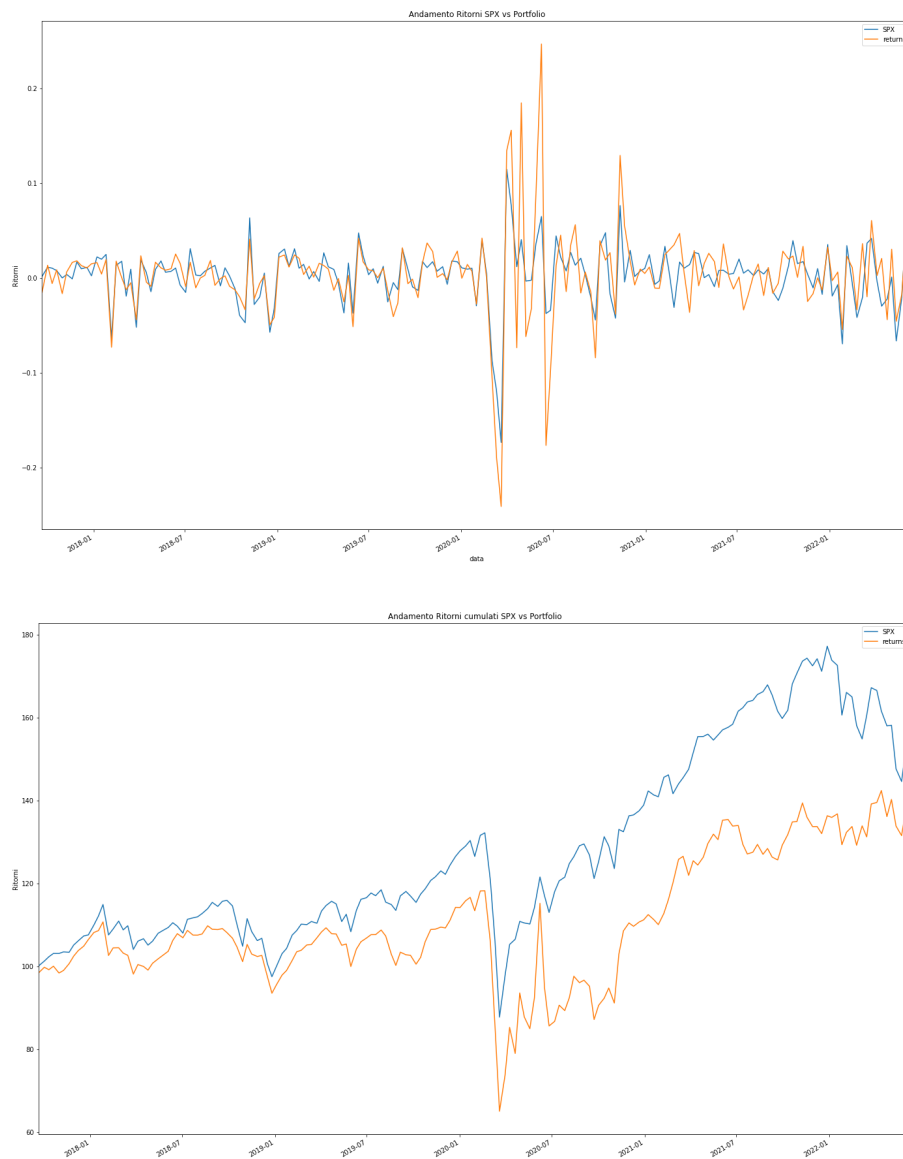
l'anno successivo overperforma il mercato con dei periodi con volatilità anche di molto maggiore rispetto a quella del mercato. Negli ultimi due anni i ritorni di S&P500 sono maggiori rispetto a quelli del portfolio; inoltre, sembra che il portfolio fatichi a seguire l'andamento dell'indice di mercato.

3.3. 10% titoli con rapporto ritorno/rischio massimo



Il portafoglio e l'indice si muovono assieme circa fino alla fine del 2019, da quel momento in poi il portafoglio ottiene dei ritorni cumulati peggiori di quelli ottenuti dall'indice.

3.4. 10% titoli con rapporto ritorno/rischio minimo



In termine di ritorni cumulati, il portafoglio con il rapporto ritorno/rischio minimo performa sempre peggio del mercato, ha ritorni inferiori.

In questa fase sono state anche realizzate delle tabelle volte a confrontare i ritorni, i rischi e il rapporto ritorni/rischio dei vari portafogli creati e dell'indice di mercato S&P500.

Di seguito si trova la tabella riguardante i portafogli creati scegliendo il 2% del totale degli stock.

	rSquareUp	rSquareDown	alphaUp	alphaDown	retrisUp	retrisDown	SPX
Expected Return	0.002881	0.000379	0.001591	0.000243	0.002560	0.002041	0.002298
Volatility	0.038497	0.024983	0.037962	0.065726	0.029711	0.055612	0.032323
Ratio	0.074828	0.015158	0.041915	0.003692	0.086154	0.036702	0.071099

Quest'altra tabella riguarda i portafogli creati scegliendo il 10% del totale degli stock.

	rSquareUp	rSquareDown	alphaUp	alphaDown	retrisUp	retrisDown	SPX
Expected Return	0.001119	0.002241	-0.000703	-0.000124	0.002022	0.002517	0.002298
Volatility	0.041375	0.030941	0.035839	0.037631	0.033485	0.048857	0.032323
Ratio	0.027043	0.072413	-0.019622	-0.003308	0.060391	0.051519	0.071099

9. Conclusioni

Dalle nostre esecuzioni emerge che la maggioranza dei portfoli generati ottengono dei ritorni minori rispetto a quelli del mercato. Ad eccezione del portfolio creato prendendo il 2% degli stock con migliori R^2 e il 2% con rapporto ritorno rischio massimo.

In quanto alla performance dei portafogli, si possono fare delle considerazioni diverse.

Difatti, come possiamo vedere dalla tabelle, rSquareUp 2% ha il ratio tra ritorno e rischio più alto tra tutti i portafogli. Notiamo come tra 2% e 10% R^2 up in quest'ultimo oltre ad aumentare la volatilità diminuisce anche il ritorno atteso.

Il portfolio generato prendendo il 2% dei titoli migliori per rapporto ritorno rischio (retrisUp) ha generato performance migliori rispetto a tutti i portfoli generati. Ha un ritorno atteso più alto rispetto al portfolio di mercato con una volatilità maggiore.

Per quanto riguarda i portfoli generati prendendo il 10% delle metriche, il portfolio con performance migliori del mercato è R^2 con i valori più bassi. Le performance sono migliori perché c'è meno volatilità, ma anche i ritorni sono minori. In questo caso l'analisi di alpha ha portato a un rapporto ritorno rischio negativo a causa dei forti ritorni negativi.